

Klasifikasi Pemasaran Produk pada Perbankan Menggunakan Metode Kuadrat Terkecil

Classification of Product Marketing in Banking Using the Least Squares Method

Yudi Aris Munandar^{1*}, Hartono²

^{1,2}Program Studi Matematika, Universitas Sanata Dharma, Yogyakarta

^{1*}yudimunandar@gmail.com

*penulis korespondensi

Abstrak

Pemasaran produk merupakan suatu hal yang sangat penting untuk menjaga kelangsungan hidup perusahaan dan juga membantu perusahaan untuk mencapai tujuannya. Permasalahan pemasaran suatu produk kepada calon nasabah dapat diformulasikan ke dalam masalah klasifikasi. Klasifikasi merupakan metode yang dapat digunakan untuk mempermudah dalam menganalisa dan mengambil suatu keputusan pada sekumpulan data. Ada beberapa metode yang dapat digunakan untuk mengklasifikasikan sekumpulan data dengan cukup baik. Pada penelitian ini, metode yang digunakan untuk mengklasifikasi adalah metode Klasifikasi Kuadrat Terkecil (*Least Square Classification*). Metode klasifikasi kuadrat terkecil adalah salah satu metode sederhana yang dapat bekerja dengan cukup baik dalam melakukan klasifikasi pemasaran pada perbankan. Secara umum, penyelesaian masalah klasifikasi kuadrat terkecil sama dengan penyelesaian model regresi yang telah ditransformasi ke dalam persamaan linear menggunakan faktorisasi QR. Pada penelitian ini, hasil klasifikasi dalam pemasaran produk perbankan dilakukan dengan data latih sebanyak 7200, dan data uji sebanyak 800 data. Hasil akurasi dari proses perhitungan dapat diperoleh melalui data uji sebesar 84%. Dari hasil pengujian yang dilakukan dapat disimpulkan bahwa metode klasifikasi kuadrat terkecil dapat mengklasifikasikan pemasaran produk pada perbankan dengan cukup baik.

Kata kunci: klasifikasi kuadrat terkecil; faktorisasi QR; pemasaran produk perbankan; akurasi

Abstract

Product marketing is a very important thing to maintain the survival of the company and also helps the company to achieve its goals. The problem of marketing a product to prospective customers can be formulated into a classification problem. Classification is a method that can be used to make it easier to analyze and make decisions on a set of data. There are several methods that can be used to classify a set of data fairly well. In this study, the method used to classify is the Least Square Classification method. The least squares classification method is a simple method that can work quite well in conducting marketing classifications in banking. In general, solving the least squares classification problem is the same as solving the regression model that has been transformed into a linear equation using QR factorization. In this study, the classification results in banking product marketing were carried out with 7200 training data, and 800 test data. The results of the accuracy of the calculation process can be obtained through test data of 84%. From the results of the tests conducted, it can be concluded that the least squares classification method can classify product marketing in banking quite well.

Keywords: least squares classification; QR factorization; banking product marketing; accuracy

1. PENDAHULUAN

Pemasaran produk merupakan suatu hal yang sangat penting untuk menjaga kelangsungan hidup perusahaan dan juga membantu perusahaan untuk mencapai tujuannya. Permasalahan pemasaran suatu produk kepada calon nasabah dapat diformulasikan ke dalam masalah klasifikasi. Klasifikasi merupakan metode yang dapat digunakan untuk mempermudah dalam menganalisa dan mengambil suatu keputusan pada sekumpulan data. Mengelompokkan atau mengklasifikasi populasi memungkinkan pihak perbankan untuk menentukan ukuran produk yang dapat dipasarkan. Sangat penting untuk mengetahui sifat atau perilaku nasabah saat ini dan dari mana nasabah baru di masa depan. Informasi ini berguna untuk membantu pihak perbankan dalam memutuskan target utama dalam penawaran yang akan dilakukan kepada nasabah. Kelompok-kelompok demografi yang mempunyai daya ketertarikan lebih tinggi akan menjadi target utama dari pihak perbankan dalam melakukan penawaran produk deposito yang disediakan. Salah satu metode untuk mengklasifikasikan pemasaran yang dapat digunakan yaitu dengan metode klasifikasi kuadrat terkecil (*least squares classification*). Klasifikasi kuadrat terkecil adalah metode sederhana yang dapat bekerja dengan baik, yang digunakan untuk membangun model klasifikasi dari sebuah kumpulan data.

Klasifikasi 2-kelas memiliki $K=2$ label kemungkinan, yaitu $y = -1$ yang berarti bernilai salah dan $+1$ bernilai benar. Klasifikasi 2-kelas adalah fungsi $\hat{f}: R^n \rightarrow \{-1, +1\}$. Model yang digunakan untuk mengklasifikasi fungsi $\hat{f}: R^n \rightarrow \{-1, +1\}$ adalah $\hat{f}(x) = \text{sign}(\tilde{f}(x))$, dengan $\tilde{f}(x) = x^T \beta + v$. Jadi model yang didapatkan adalah:

$$\hat{f}(x) = \text{sign}(x^T \beta + v).$$

dengan $\text{sign}(a) = +1$ untuk $a \geq 0$ dan $\text{sign}(a) = -1$ untuk $a < 0$. Fungsi $\text{sign}(a)$ inilah yang disebut fungsi pengklasifikasi.

2. METODE

2.1 Fungsi Affine

Fungsi afin adalah fungsi linear yang ditambah sebuah konstanta (1). Sebuah fungsi $f: R^n \rightarrow R$ adalah fungsi afin, jika dan hanya jika dapat ditulis sebagai $f(x) = a^T x + b$. Untuk $a \in R^n$ dan $b \in R$, dengan a adalah vektor berdimensi n dan b adalah sebuah konstanta.

2.2 Substitusi Balik (*Back Substitution*)

Algoritma substitusi balik (*back substitution*), biasa digunakan untuk menyelesaikan suatu persamaan linear, $Ax = b$ [2]. Persamaan $Ax = b$ dapat ditulis sebagai berikut:

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \dots + A_{1,n}x_{n-1} + A_{1n}x_n &= b_1 \\ &\vdots \\ A_{n-2,n-2}x_{n-2} + A_{n-2,n-1}x_{n-1} + A_{n-2,n}x_n &= b_{n-2} \\ A_{n-1,n-1}x_{n-1} + A_{n-1,n}x_n &= b_{n-1} \\ A_{n,n}x_n &= b_n \end{aligned}$$

Dari persamaan terakhir, dapat ditemukan bahwa $x_n = b_n/A_{nn}$. Dari hasil yang didapatkan, x_n dapat di substitusikan pada persamaan di atasnya. Lalu dapat dilanjutkan dengan cara yang sama, yaitu dengan mensubstitusikan hasil yang didapatkan untuk menemukan $x_{n-2}, x_{n-3}, \dots, x_1$.

2.3 Faktorisasi QR

Faktorisasi QR adalah bentuk transformasi dari suatu matrik A yang berukuran $n \times k$ menjadi,

$$A = QR \quad (1)$$

Bentuk $A = QR$ disebut sebagai faktorisasi QR dari matrik A [3], matrik Q berukuran $n \times k$ memiliki vector kolom ortonormal, dan matrik R berukuran $k \times k$ adalah matrik segitiga atas dengan elemen diagonal positif.

2.4 Pemasangan Data Kuadrat Terkecil (*Least Square Data Fitting*)

Least Square Data Fitting digunakan untuk memilih model parameter $\theta_1, \dots, \theta_p$, dengan nilai *Root Mean Square Error* (RMSE) yang minimum [1]. Untuk menyelesaikan permasalahan dalam mencari nilai θ , yaitu sama dengan menyelesaikan persamaan linear $\hat{y} = A\theta$ dengan metode QR faktorisasi dan substitusi balik.

$$\begin{aligned} \hat{y} &= \hat{f}(x) \\ &= \theta_1 f_1(x) + \dots + \theta_p f_p(x) \\ &= A\theta \end{aligned}$$

2.5 Model Regresi

Model regresi adalah bentuk umum dari fungsi afin. Fungsi afin dari x yang sangat umum digunakan adalah sebagai berikut:

$$\hat{y} = X^T \beta + v, \quad (2)$$

Vektor β disebut vector bobot, atau vektor koefisien, dan skalar v disebut intersep dalam model regresi [4]. Model regresi dapat direpresentasikan menggunakan model data *fitting* dengan basis fungsi $f_1(x) = 1$, dan $f_i(x) = x_{i-1}$ dimana $i = 2, \dots, n + 1$. Jadi model regresi yang didapat adalah:

$$\begin{aligned} \hat{y} &= X^T \theta_i + \theta_1 = [1 \quad X^T] \theta, \quad i = 2, \dots, n + 1 \\ \hat{y} &= [1 \quad X^T] \theta = A\theta. \end{aligned} \quad (3)$$

Dengan $\beta = \theta_i$ dan $v = \theta_1$.

2.6 Klasifikasi Kuadrat Terkecil (KKT)

Pengklasifikasian kuadrat terkecil sering menggunakan model regresi, dengan $\tilde{f}(x) = X^T \theta_i + \theta_1$ [5]. Jadi model yang didapatkan:

$$\begin{aligned} \hat{f}(x) &= \text{sign}(\tilde{f}(x)) \\ \hat{f}(x) &= \text{sign}(X^T \theta_i + \theta_1) \end{aligned} \quad (4)$$

Dimana,

$$\text{sign}(a) = \begin{cases} +1, & \text{untuk } a \geq 0 \\ -1, & \text{untuk } a < 0 \end{cases}$$

2.7 Desain Alat Uji

Pada bagian ini akan dijelaskan perancangan sistem dari awal proses data hingga mendapatkan hasil akhir. Berikut tabel perencanaan atau gambaran sistem secara umum:



3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dalam melakukan penelitian penulis menggunakan sebuah perangkat lunak *python*, dan *microsoft excel* dioperasikan pada perangkat keras komputer laptop dengan spesifikasi Core i3, RAM 4 GB, dan system windows 10. Pada penelitian ini data yang digunakan adalah data promosi pemasaran dari lembaga perbankan Portugis. Dimana data pada penelitian ini adalah data sekunder dimana data diambil pada tanggal 12 juni 2022, yang telah tersedia pada situs *web uci machine learning*: <https://archive.ics.uci.edu/ml/datasets/bank+marketing>[6]. Data yang disediakan pada *web uci machine learning* terdapat dua variasi data, data pertama sebesar 45211, dan data kedua sebesar 4521, dataset terdiri dari 16 atribut dan satu atribut prediksi(*output*). Pada penelitian ini data yang digunakan sebesar 4521 dataset dengan 16 atribut dan satu atribut prediksi(*output*). Berikut adalah bentuk dataset yang telah diambil beserta keterangannya (12):

Tabel 1. Data Set Nasabah

Nama Atribut	Tipe Data
y	No, Yes
age	19-87thn
job	admin, unknown, unemployed, management, housemaid, entrepreneur, student, blue-collar, self-employed, retired, technician, services
marital	married, divorced, single
education	Unknown, secondary, primary, tertiary
default	No, Yes
balance	-3.313-71.188
housing	No, Yes
loan	No, Yes
contact	Unknown, telephone, cellular
day	1-31
month	Januari-desember
duration	0-3025s
campaign	1-50
pdays	-1 – 871
previous	0-25
poutcome	Unknown, other, failure, success

3.2 Preprocessing Data

3.2.1 Data Cleansing

Data *cleansing* merupakan proses membersihkan data. Data yang dibersihkan tersebut adalah data yang tidak akurat, tidak lengkap, dan data yang terpencil atau *outliers* (14). Penghapusan nilai *outlier* pada atribut bertujuan agar variasi nilai tidak jauh berbeda. Penghapusan *outlier* akan dilakukan dengan metode *Interquartile Range* (IQR) [7]. Untuk mencari nilai *outlier* dapat menggunakan formula sebagai berikut:

$$\begin{aligned}
 IQR &= Q_3 - Q_1 \\
 \text{nilai minimum} &= Q_1 - 1.5 * IQR \\
 \text{nilai maksimum} &= Q_1 + 1.5 * IQR
 \end{aligned}$$

3.2.2 Normalisasi Data

Normalisasi adalah proses transformasi beberapa variabel dengan menempatkan pada interval [0,1]. Normalisasi ini bertujuan agar *range* (jangkauan) data tiap atribut menjadi seragam atau tidak berbeda jauh. Metode normalisasi yang akan dipakai dalam penelitian ini yaitu metode *min-max* (8). Berikut adalah formula dalam normalisasi:

$$X_{new} = \frac{X_i - X_{minimum}}{X_{maksimum} - X_{minimum}}$$

Atribut yang akan melalui proses normalisasi adalah atribut yang berbentuk numerik. Berikut adalah tabel nilai minimum dan maksimum dari atribut yang akan di normalisasi:

Tabel 2. Atribut Numerik

Atribut	Nilai Minimum	Nilai Maksimum
<i>Age</i>	19	87
<i>Balance</i>	-3313	71188
<i>Duration</i>	0	3025
<i>Day</i>	1	31
<i>campaign</i>	1	50

3.2.3 Transformasi Variabel Dummy

Untuk data kategori atau atribut yang berisikan data *nonnumerik* akan di transformasi menggunakan variabel *dummy*. Variabel *dummy* digunakan untuk merepresentasikan variabel yang bersifat kualitatif dengan nilai 0 dan 1 (9). Atribut yang akan ditransformasi yaitu atribut *job*, *marital*, *education*, *contact*, *month*, dan *poutcome*.

3.2.4 Imbalanced Data (Ketidakseimbangan Data)

Imbalance data merupakan suatu keadaan dimana distribusi kelas tidak seimbang di setiap kelasnya (10). Metode untuk mengatasi *imbalance* yang digunakan yaitu, *random oversampling* (ROS). *Random oversampling* adalah teknik menduplikasi sampel secara acak pada kelas minoritas yang bertujuan agar jumlah data sampel sama dengan jumlah data sampel mayoritas, lalu membentuk sebuah dataset baru (11). Dalam penelitian ini jumlah sampel data kelas dengan label 0 sebesar 4000 dan data sampel kelas dengan label 1 sebesar 521, sehingga data sampel pada label 1 secara acak diduplikasi sehingga jumlahnya sama dengan data sampel label 0. Berikut adalah hasil dari proses *imbalanced*:

```
from imblearn.over_sampling import RandomOverSampler

ros = RandomOverSampler(sampling_strategy= 1)
X_res, y_res = ros.fit_resample(x, y)
yres = y_res
yres.value_counts()

0    4000
1    4000
..     ..
```

Figure 1 Code Proses Imbalance Pada Phyton

3.2.5 Pembagian Data Latih Dan Data Uji

Pembagian untuk data latih sebesar 90% dan untuk data uji sebesar 10%. Pembagian tersebut menghasilkan data latih sebesar 7200, dan data uji sebesar 800. Data latih digunakan untuk melatih algoritma klasifikasi, kemudian data uji digunakan untuk menguji algoritma/model yang telah dilatih (15). Selanjutnya akan dilakukan *K-fold cross validation* dengan $K = 10$ pada data latih. *K-fold cross validation* bertujuan untuk mencari model terbaik dengan RMSE yang paling minimum untuk digunakan dalam model klasifikasi (13). RMSE yang diperoleh dari *cross validation* pada percobaan pertama sampai percobaan terakhir atau ke-10 adalah sebagai berikut:

Tabel 3. RMSE Proses Cross-Validation

RMSE TEST	
1	0.27
2	0.28
3	0.29
4	0.27
5	0.28
6	0.28
7	0.33
8	0.33
9	0.32
10	0.32

$$\begin{aligned}
 RMSCVE &= \sqrt{\frac{e_1^2 + \dots + e_n^2}{n}} \\
 &= \sqrt{\frac{(0,27)^2 + (0,28)^2 + (0,29)^2 + (0,27)^2 + (0,28)^2 + (0,28)^2 + (0,33)^2 + (0,33)^2 + (0,32)^2 + (0,32)^2}{10}} \\
 &= 0,29
 \end{aligned}$$

Dari *RMSE* yang diperoleh dapat diketahui *Root Mean Square Cross-Validation Error (RMSCVE)* sebesar 0,29 atau sebesar 29%. Sehingga pada percobaan ke-3 mempunyai eror yang paling minimum daripada percoaan yang lain.

3.3 Hasil Klasifikasi Kuadrat Terkecil

Pada proses percobaan di atas telah diperoleh model yang baik dengan *RMSE* yang paling minimum. Model yang diperoleh akan digunakan untuk mengklasifikasikan data set pada data uji. Hasil klasifikasi dari data uji disajikan dalam bentuk matrik konfusi. Berikut adalah hasil klasifikasi yang diperoleh.

Tabel 1. Matrix Confusion

Keluaran	Prediksi		Total
	$\hat{y} = -1$	$\hat{y} = +1$	
$y = -1$	590	116	706
$y = +1$	15	79	94
Semua	605	195	800

Hasil dari klasifikasi 800 data uji ditampilkan pada matrix konfusi di atas, hasil dari klasifikasi terdapat 590 *true negatif*, 79 *true positif*, 15 *false positif*, 116 *false negatif*. Dari hasil yang didapat dapat dihitung kinerja matrik konfusi sebagai berikut:

- Tingkat kesalahan = $\frac{15+116}{800} = 0.16 = 16\%$ (artinya jumlah keseluruhan kesalahan dalam mendeteksi data BENAR ($y = +1$) dan data SALAH ($y = -1$) sebesar 16%)
- True negatif rate = $\frac{590}{706} = 0.835 = 83,5\%$ (artinya dapat mendeteksi data BENAR ($y = +1$) dengan benar sebesar 83,5%)

- False positive rate = $\frac{15}{94} = 0.16 = 16\%$ (artinya data SALAH ($y = -1$) yang dideteksi sebagai data BENAR ($y = +1$) sebesar 16%)
- True positif rate = $\frac{79}{94} = 0.84 = 84\%$ (artinya dapat mendeteksi data SALAH ($y = -1$) dengan benar sebesar 84%)
- False negatif rate = $\frac{116}{706} = 0.16 = 16\%$ (artinya data BENAR ($y = +1$) yang dideteksi sebagai data SALAH $y = -1$) sebesar 16%)
- Precision = $\frac{590}{590+15} = 0.97 = 97\%$

4. KESIMPULAN DAN SARAN

Dari hasil penelitian pada klasifikasi pemasaran produk pada perbankan menggunakan metode klasifikasi kuadrat terkecil dapat disimpulkan bahwa, metode klasifikasi kuadrat terkecil dapat digunakan sebagai alat klasifikasi pemasaran pada perbankan dengan cukup baik. Dari data penelitian sebanyak 8000 dataset dan 40 atribut dengan pembagian data sebanyak 7200 data latih, dan 800 data uji didapatkan akurasi sebesar 84%. Saran untuk penelitian selanjutnya dapat menggunakan metode PCA, untuk mereduksi dimensi dari atribut pada dataset.

DAFTAR PUSTAKA

- (1) Boyd, Stephen and Lieven Vandenberghe. (2018). *Introduction to Applied Linear Algebra: Vector, Matrices, and Least Squares*. London: Cambridge University.
- (2) Anton, Howard & Chris Rorres. (2014). *Elementary Linear Algebra (11ed)*. Canada: Anton Textbooks, Inc.
- (3) Yanita dkk. (2020). "Menghitung Determinan Matriks Blok Menggunakan Ekspansi Laplace dan Komplemen SCHUR". *Jurnal Matematika UNAND*. Vol. 9, No. 2, Hal. 138-145.
- (4) Wooldridge, Jeffrey M. (2009). *Introductory Econometrics A Modern Approach*. Canada: Nelson Education, Ltd.
- (5) Leung, Jessica & Dmytro Matsypura. (2019). *Python Language Companion to "introduction to Applied Linear Algebra: Vector, Matrices, and Least Squaer"*. Sidney.
- (6) Moro dkk. (2014). "A Data-Driven Approach to Predict the Success of Bank Telemarketing". Elsevier. <https://archive.ics.uci.edu/ml/datasets/bank+marketing>.
- (7) Junaidi. (2015). "Deskripsi Data Melalui Box-Plot". Jambi: Universitas Jambi.
- (8) Adiwijaya dkk. (2017). "Analisis Churn Prediction pada Data Pelanggan PT.Telekomunikasi dengan Logistic Regression dan Underbagging". *e-Proceeding of Engineering*. Vol. 4, no. 2
- (9) Artaya, I Putu. (2019). "Regresi Linear Berganda Metode Dummy". Surabaya: Universitas Narotama.
- (10) Ali dkk. (2015). "Classification With Class Imbalance Problem: A Review". *Int J Adv. Soft Compu. Appl*. Vol. 7, no. 3.
- (11) Sukron, Akhmad & Agus Subekti. (2018). "Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit". *Jurnal Informatika* Vol.5, no.2.
- (12) Fauzi, Ahmad. (2021). "Analisis Data Bank Direct Marketing dengan Perbandingan Klasifikasi Data Mining Berbasis Optimize Selection (Evolutionary)". *Jurnal Informatika Universitas Pamulang*, Vol. 6, no. 1.

- (13) Rhomadhona, Herfia & Jaka Permadi. (2019). “Klasifikasi Berita Kriminal Menggunakan Naïve Classifier (NBC) dengan Pengujian *K-Fold Cross Validation*”. *Jurnal Sains dan Informatika*. Vol. 5, no. 2.
- (14) Lazuardy, Ainayya Ghassani & Hari Setiaji. (2019). “Data Cleanssing Pada Data Rumah Sakit”. *Proceeding SINTAK*.
- (15) Saifudin, Aries. (2017). “Metode Data Mining Untuk Seleksi Calon Mahasiswa Pada Penerimaan Mahasiswa Baru Di Universitas Pamulang”. *Jurnal Teknologi*. Vol. 10, no. 1.